RESEARCH ARTICLE

10.1002/2016WR018692

Key Points:

- A methodology is developed to identify (in)consistency in rating curve data
- The methodology is successfully validated with both measured and synthetic data series
- For all validation cases, the methodology turns out robust, showing little dependency of decisions

Correspondence to:

K. Van Eerdenbrugh,
katrien.vaneerdenbrugh@ugent.be

Citation:

Van Eerdenbrugh, K., S. Van Hoey, and N. E. C. Verhoest (2016), Identification of temporal consistency in rating curve data: Bidirectional Reach (BReach), *Water Resour. Res.*, 52, 6277–6296, doi:10.1002/2016WR018692.

Received 28 JAN 2016

Accepted 17 JUL 2016

Accepted article online 22 JUL 2016

Published online 18 AUG 2016

Identification of temporal consistency in rating curve data: Bidirectional Reach (BReach)

Katrien Van Eerdenbrugh¹, Stijn Van Hoey², and Niko E. C. Verhoest¹

¹Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium, ²BIOMATH, Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Ghent, Belgium

Abstract In this paper, a methodology is developed to identify consistency of rating curve data based on a quality analysis of model results. This methodology, called Bidirectional Reach (BReach), evaluates results of a rating curve model with randomly sampled parameter sets in each observation. The combination of a parameter set and an observation is classified as nonacceptable if the deviation between the accompanying model result and the measurement exceeds observational uncertainty. Based on this classification, conditions for satisfactory behavior of a model in a sequence of observations are defined. Subsequently, a parameter set is evaluated in a data point by assessing the span for which it behaves satisfactory in the direction of the previous (or following) chronologically sorted observations. This is repeated for all sampled parameter sets and results are aggregated by indicating the endpoint of the largest span, called the maximum left (right) reach. This temporal reach should not be confused with a spatial reach (indicating a part of a river). The same procedure is followed for each data point and for different definitions of satisfactory behavior. Results of this analysis enable the detection of changes in data consistency. The methodology is validated with observed data and various synthetic stage-discharge data sets and proves to be a robust technique to investigate temporal consistency of rating curve data. It provides satisfying results despite of low data availability, errors in the estimated observational uncertainty, and a rating curve model that is known to cover only a limited part of the observations.

1. Introduction

Discharge estimation through rating curves is a widespread technique in the field of hydrologic monitoring. A rating curve relates discharge to stage measurements in a river section. This rating curve, which is generally derived from a limited number of flow measurements, then allows to transform high-frequency stage records into discharge values. This is an effective method to avoid the installation of expensive discharge measurement stations that require intensive maintenance.

When calibrating rating curves, data consistency is a critical issue. Especially for locations that are subjected to changes in river geometry or flow conditions (e.g., caused by weed growth), it is crucial to detect and describe both rapid and gradual changes in rating curve behavior. In operational hydrology, a commonly used practice for the assessment of these changes is to examine the difference between the current rating curve and new observations [Reitan and Petersen-Øverleir, 2011]. Consecutive and systematic deviations that exceed observational uncertainty are an indication for a transition in data consistency. The success of this approach depends upon measurement frequency and accuracy, knowledge of the site-specific conditions, and experience of the analyst [Reitan and Petersen-Øverleir, 2011]. Moreover, it is hard to pinpoint gradual changes in river cross section or in roughness objectively with this technique.

Data consistency in rating curve observations has been studied by several authors and with a diversity of methods. A rather pragmatic approach was applied by McMillan *et al.* [2010], who selected the 0.5 year return period discharge as the threshold to define a major event. These thresholds are used to identify periods of stationarity between major flood events and thus the assumption is made that floods are the only causes of change in rating curve behavior. Westerberg *et al.* [2011] implemented a weighted fuzzy regression technique with a moving time window to detect temporal variability of stage-discharge relationships. Comparably, in the field of hydrologic modeling, a dynamic identifiability analysis is used to identify time dependency of parameters [Wagner *et al.*, 2003; Van Hoey *et al.*, 2014]. These methods have the

disadvantage that results depend on the choice of an appropriate window size for the moving window. This dependency is expected to gain importance in case of a limited data availability or an irregular time spread of observations. Both of these data restrictions often occur in rating curve data series and hence limit the applicability of these methods. Another approach was proposed by *Reitan and Petersen-Øverleir* [2011], who applied a hierarchical Bayesian analysis that makes use of the stage-discharge observations. The rating curve parameters are viewed as continuous time stochastic processes and the final goal of this analysis is to assess parameters that are used to describe a time evolution of these rating curve parameters. The method combines results of different submodels that simulate this time evolution, each considering one or more rating curve parameters to be constant. The preferred submodel is selected using Bayesian model hypothesis testing. This selection is site-specific and sometimes alters with small changes in the prior information that is used. At the same time, *Jalbert et al.* [2011] investigated nonstationarity of stage-discharge relations based on a variographic analysis. Within predefined validity periods, the authors consider both an initial rating curve error and an aging error, the latter caused by the increasing risk of a change in the river bed over time. To overcome the predefined validity periods in this method, *Morlot et al.* [2014] expanded it with two preliminary steps. The first consists of a chronological segmentation of the observations by clustering the differences between a mean rating curve and measurements. In a second step, an in-depth analysis is performed to select hydraulic analogs of each gauging within a consistent data period. Although observational uncertainties are considered in the latter step, the first and hence defining segmentation of consistent data periods does not account for them.

All above mentioned methods for the detection of temporary or permanent changes of the reference hydraulic regime involve assumptions or decisions that more or less influence eventual results. Moreover, the majority of them depart from the definitive choice of a rating curve model and include an assessment of its parameters distribution. However, a preliminary approach to detect data consistency prior to an in-depth analysis as in *Morlot et al.* [2014] forms a logical sequence. Robustness of this first step, i.e., little dependency of assumptions and choices to be made is then a crucial success factor.

The Bidirectional Reach (BReach) methodology presented in this paper is established to provide this preliminary identification of data consistency through an analysis of realizations of a rating curve model. Crucial in this approach is to start from a straightforward definition of a consistent period. Similarly as in operational hydrology, a period is considered to be consistent if no consecutive and systematic deviations from a current situation occur that exceed observational uncertainty. This definition requires the assessment of

1. Observational uncertainty.
2. A current situation.
3. The consecutive and systematic character of nonacceptable deviations.

Unless more local information is available, literature values of uncertainties on stage and discharge measurements can be used for the first item. In this paper, a model (i.e., the combination of a rating curve model and a parameter set) and its accompanying results are classified as acceptable for a gauging point if these results fit in a rectangular acceptance zone that is enclosed by the 95% uncertainty boundaries of the stage and discharge measurements, respectively.

The assessment of a current situation is less evident. In this research, this is done by assessing the capability of a model to describe a subset of the data. In previous studies, several authors already used rating curve models that only describe a subset of the data. In *McMillan et al.* [2010], these subsets are based on the occurrence of peak discharges above a defined threshold. *Krueger et al.* [2010] fit rating curve parameters on different data subsets and evaluate each parameter set with its performance for the remaining data and *Morlot et al.* [2014] work with a selection of data measured under the same hydraulic conditions. In *McMillan and Westerberg* [2015], a likelihood weight for uncertainty assessment is proposed that uses the size of the subset that can be described by one rating curve as an input. Inspired by the latter, the BReach methodology introduces the maximum left and right reach of a model in each data point of a chronologically sorted data series. This maximum reach of a model in a data point is assessed in both the direction of the previous data points (left) and of the following data points (right) and is the index (and accompanying data point) beyond which none of the sampled parameter sets both are satisfactory and result in an acceptable deviation. Additionally to the above mentioned definition of acceptable deviations, a definition of satisfactory behavior of a model and its results in a sequence of data points is thus required. In this paper, the

percentage of data points that are allowed to have a nonacceptable model result defines satisfactory behavior. This condition is called the degree of tolerance.

To avoid confusion between a temporal reach (as defined in this paper) and a spatial reach (often used to indicate a part of a river), the latter is consequently referred to as “river reach.”

Combining the resulting maximum left and right reaches for different degrees of tolerance leads to the third requirement in the definition of consistent periods, the assessment of the consecutive and systematic character of nonacceptable deviations. This is illustrated below with an example.

The assessment of the maximum reaches is illustrated in Figure 1 for a set of 276 gauging points that are approximated using a power law with three parameters as a rating curve model:

$$Q=c.(h-h_0)^n \quad (1)$$

where Q is the discharge [m^3/s], c is a scale coefficient [m^3/s], h is the stage [m], h_0 is a location parameter [m] expressing the stage of zero flow, and n is an exponent [–] whose value depends on the type and the shape of the considered cross section(s).

First, a degree of tolerance of 0% is assumed, which implies that model results must fit in the acceptance zone of all data points to be satisfactory. All data points are sorted chronologically and the index that indicates the position of a data point within the data series is used to refer to it. In a first example, one set of parameters is considered that results from the minimization of the root-mean-square error (RMSE) in all data points. It should be noted that the choice to illustrate the methodology using this model, further referred to as “the RMSE optimized model,” is arbitrary and does not influence final results. For this model, the left and right reaches are assessed in Figure 1a. For each index in the x axis, the colored area shows the span between the index of the left reach (under the bisector) and the right reach (above the bisector). This is illustrated in the detailed part of the figure for the data point with index 125. This vertical distance between the left and right reach indicates the maximum amount of data points around the index of interest that can be described by the model under the prevailing degree of tolerance. Each discontinuity in the left or right reach indicates the presence of a gauging point for which the RMSE optimized model generates unacceptable results. The figure shows that no series of consecutive acceptable model results is longer than seven data points. However, it is likely that other parameter sets with higher RMSE values exist that reach further. Therefore, all parameter sets that were sampled for the RMSE optimization are included in a second analysis. For the three parameters, ranges for these samples are based on physical considerations (section 2.1.2). For each of these samples, the left and right reach in each data point are assessed similarly as in Figure 1a. Subsequently, the outermost indices of all these reaches in a point are selected as the maximum left and right reach and are shown in a BReach plot (Figure 1b). This plot confirms that in nearly all data points, at least one parameter set exists that reaches further than the RMSE optimized model. When analyzing this plot, it is important to notice that the distances between the bisector and respectively the maximum left reach and the maximum right reach in a point are not necessarily caused by the same parameter set(s). Again, discontinuities in one of the maximum reaches indicate the presence of a gauging point for which the models that realize this reach have unacceptable results.

As the degree of tolerance is very stringent, the discontinuities in the BReach plot provide information on the diversity of the data points that is related with one of the following causes: (1) the occurrence of a higher observational error than estimated for the definition of acceptable results, (2) model deficiency in certain ranges of the investigated variables, and (3) data inconsistency. It can be expected that discontinuities caused by (1) and (2) tend to be singularities in the data series. Even if a few consecutive stage-discharge data are measured in a short time span, there is a (often wanted) large variation in height range and flow situations of the stage-discharge data in time. Therefore, observational errors and model structural errors in these data points can be considered to occur randomly and independent of errors in neighboring points. On the contrary, if a discontinuity is caused by data inconsistency, it can be expected that the following data points will show a comparable change in behavior toward the previous data points. For that reason, the same analysis is performed with more relaxed degrees of tolerance that allow model results to be unacceptable in 5%, 10%, 20%, and 40% of the data points. For these more relaxed degrees of tolerance, only a limited subset of the observations is required to have acceptable model results for the assessment of both reaches. Therefore, limits in the maximum reach due to causes (1) and (2) will disappear from the plots

Figure 1. BReach plot of a synthetic data series with a created abrupt change in data consistency at index 201 for a degree of tolerance of (a) 0% and only one parameter set in the analysis, with indication of maximum left reach (◀), maximum right reach (▶), and analyzed index (●) for the data point with index 125 (b) 0% indicating possible last points of a first consistent period (◊) and possible first points of a second consistent period (♦) (c) 40% with two indicative transition points (*). In all subplots, for each index in the x axis the colored area indicates the span between the index of the maximum left reach (under the bisector) and the maximum right reach (above the bisector).

Figure 2. Combined BReach plots of synthetic data series with a created abrupt change in data consistency (a) at index 201, (b) at index 48, (c) at index 103. In all subplots, for each index in the x axis the gray area indicates the span between the index of the maximum left reach (under the bisector) and the maximum right reach (above the bisector). Each gray tint represents a different degree of tolerance.

and those caused by data inconsistency will be stressed. This is confirmed in Figure 1c, showing a BReach plot for a degree of tolerance of 40%. In this plot, only one of the discontinuities of Figure 1b remains visible. In Figure 2a, a combined BReach plot is shown with results of all different degrees of tolerance. In this combined plot, a change in data consistency is visible for all degrees of tolerance at the data point with index 201. This corresponds with the characteristics of the data series that is used for this illustrative example, which is a synthetic data series with a created transition in data consistency at index 201.

As the different degrees of tolerance provide complementary information, a wide range of degrees of tolerance is combined when performing a BReach analysis. These results are used for the last requirement in the definition of consistent periods, i.e., the assessment of the consecutive and systematic character of nonacceptable deviations.

The different steps that constitute the BReach methodology are discussed more exhaustively in the next section. The methodology is validated with measured data and various synthetic data sets to test (1) the basic capability of the methodology to detect changes in data consistency, robustness when (2) data availability or (3) observational uncertainties alter and (4) effects of known model structural errors on results.

2. Methods

2.1. BReach Methodology: Description and Application

BReach (Bidirectional Reach) is a methodology to reveal patterns in consistency of rating curve data based on a quality analysis of model results. The consecutive steps in this methodology are:

1. Selection of a model structure for the analysis.
2. Sampling of the parameter space.
3. Assessment of acceptable model results.
4. Assessment of different degrees of tolerance.
5. Assessment of the bidirectional reach for all degrees of tolerance.
6. Identification of consistent data periods.

Each of these steps is discussed in the following subsections.

2.1.1. Selection of a Model Structure for the Analysis

A first step in the BReach methodology is the choice of a rating curve model that appropriately describes the relation between discharge and stage for an important part of the measured range. In practical applications, rating curves are often expressed as a power law (equation (1)). This approach is allowed when the flow is controlled locally and determined by both the (regular) geometry and the roughness of a narrowing construction or section in the river. If the flow is determined by the geometry, slope, and roughness of a more extended river reach, this equation can be derived from uniform flow theory in specific situations. First, flow conditions must be uniform and no lateral inflow occurs in the river reach. Second, the cross section is wide, has a regular shape and is prismatic along the river reach. Moreover, a resistance coefficient (e.g., Manning or Chézy coefficient) is assumed constant for all stages [Schmidt and Yen, 2001; Petersen-

Øverleir and Reitan, 2009]. Nevertheless, a power law can fit empirically for other situations in which some of these (stringent) assumptions are violated. However, very irregular cross sections, highly dynamic flow conditions or important backwater effects can alter the behavior of stage-discharge data fundamentally, inducing large uncertainties in discharges simulated with a power law [e.g., Petersen-Øverleir and Reitan, 2005, 2009; Petersen-Øverleir, 2006; Dottori et al., 2009].

Despite of these restrictions, a simple power law is used to validate the BReach methodology both with measured data and with synthetic data that contain an interruption in data consistency. This choice is justified for most of the cases investigated in this paper. Nevertheless, in some of the cases these assumptions are deliberately violated. Hence, an appropriate test of the methodology's robustness with relation to model structural errors is provided.

2.1.2. Sampling of the Parameter Space

A strategy for sampling the parameters of the rating curve model has to be decided on. In the BReach methodology, this strategy determines the parameter sets that will be investigated, and is thus implicitly represented in results. An appropriate choice of a sampling strategy is thus important.

For each set of data points in this paper, the sampling strategy for the parameters of the rating curve model is the same. Although in some studies, values for the location parameter h_0 are sampled from a very wide interval [e.g., Moyeed and Clarke, 2005], many authors limit its sample range to values that approach the (mean) river bed [e.g., Pappenberger et al., 2006; Le Coz et al., 2014; McMillan and Westerberg, 2015]. In this paper, h_0 is sampled from the interval $[h_{bed} - 40 \text{ cm}, h_{min} - 15 \text{ cm}]$, where h_{bed} is the lowest bed level of the available reliable cross-section measurements that accompany a measured data set. For synthetic data sets, it is the local height of the river bed in the hydraulic model. The value of h_{min} is the lowest water level in the data series. Based on theoretical values for n , that depend on the type and shape of the cross section [Le Coz et al., 2014], samples of n are limited to the interval $[0.5, 3.5]$. The sampling interval for the coefficient c of the power law model is based on the outermost values obtained when applying the power law function for all data points with the upper and lower limits for h_0 and n . The resulting lowest value for c is halved to form the lower limit while the highest value is doubled to determine the upper limit. It is a deliberate choice to truncate the sample intervals and thus to limit parameter values within a physically realistic order of magnitude. This choice avoids the use of parameter sets that may fit well for the measured data, but perform poorly when extrapolating the rating curves toward higher discharges [Le Coz et al., 2014].

The simple structure of the power law model (that allows for abundant calculations within a relatively short time period) and the limited sampling intervals permit an intensive sampling that covers the complete parameter space. As no preference is given to certain values of h_0 and n , they are both sampled from a uniform distribution. Parameter c is sampled from a uniform distribution after log-transformation in order to acquire a more dense sampling for smaller values. A Latin Hypercube sampling is performed on the three-dimensional parameter space and 1.3×10^6 times samples are taken.

2.1.3. Assessment of Acceptable Model Results

In a next step, the generated model outputs are evaluated with respect to each available gauging point and classified as acceptable or nonacceptable. This requires the definition of a quality level that enables this classification. In this paper, a crisp discrimination is applied that accounts for uncertainties on both measured stages and discharges. A model, i.e., the power law simulated with one of the sampled parameter sets, and its accompanying results are acceptable in a data point if these results fit within the rectangular zone that is enclosed by the 95% uncertainty boundaries of the stage and the discharge measurement. This corresponds with the approach in several studies that only accept model results within the uncertainty bounds of the data used for model evaluation [e.g., Beven, 2006; Pappenberger et al., 2006; Blazkova and Beven, 2009]. Based on observational uncertainties of discharge and stage reported in the literature [e.g., Pelletier, 1988; Pappenberger et al., 2006; McMillan et al., 2010, 2012], 95% boundaries are assumed to be $\pm 1 \text{ cm}$ for stage and $\pm 6\%$ for discharge. This estimation of uncertainty could be commented as rather low, especially for the highest and lowest flows. Therefore, the effect of misjudging these measurement uncertainties on BReach results will be investigated (section 2.3.3).

The result of this step is a binary matrix with classification results for each parameter set and each data point. This binary character implies the possible presence of false rejections (type II errors) of model results for those data points that have a larger measurement error than covered by the acceptance zone.

Figure 3. (a) The River Demer situated in the Scheldt river basin. (b) The study area as covered by the MIKE 11 model.

2.1.4. Assessment of Different Degrees of Tolerance

Next, different degrees of tolerance have to be defined. A degree of tolerance is the condition that defines satisfactory behavior of a model and its results for a sequence of data points. In the cases applied in this study, a degree of tolerance is defined by the percentage of points within a data series that can have nonacceptable model results. It is important to note that these degrees of tolerance express satisfactory behavior as defined within the context of a BReach analysis, but do not necessarily correspond with satisfactory behavior in terms of final model performance. In an in-depth analysis of consistent data periods that follows a BReach analysis, more severe conditions for satisfactory behavior can be imposed to model results.

These degrees of tolerance are comparable with the limits of acceptability as defined in the Generalized Likelihood Uncertainty Estimation (GLUE) [Beven, 2006; Beven and Binley, 1992] that

serve to distinguish between behavioral and nonbehavioral models. However, the input for BReach is not limited to one fixed degree of tolerance. It requires a range of degrees, varying from highly demanding to extremely flexible. The flexible degrees will avoid the disadvantage of possible type II errors, but are also more likely to include type I errors (false acceptances). Each of these degrees offers thus additional insight for the interpretation of results.

Also the capability of the rating curve model to describe the complete range of investigated data points will influence the range of degrees of tolerance to be chosen. If the model is known to have only a limited scope, it is necessary to include sufficiently large degrees of tolerance. On the contrary, a very large degree of tolerance can completely conceal the variability in a data set. It is hence recommended to assign a wide range of degrees of tolerance for all applications. For the power law rating curves, that can contain model errors caused by governing flow conditions and varying cross-section properties, degrees of 0%, 5%, 10%, 20%, and 40% are used for all cases.

2.1.5. Assessment of the Bidirectional Reach for all Degrees of Tolerance

For all degrees of tolerance, the bidirectional reach of each of the data points has to be assessed based on the classification matrix of the chronologically sorted data series. First, a degree of tolerance is assumed. Subsequently, for each data point the temporal span for which a parameter set behaves satisfactory (according to this degree of tolerance) is assessed along this chronological data series both in the direction of the previous data points (left span) and of the following data points (right span). Within these two spans, the index of the furthest observation with an acceptable result is called the left or right reach, depending of the direction of the investigation. When aggregating the reaches in a data point for all parameters sets, points of interest are the outermost left and right reaches. These outermost reaches represent indices beyond which none of the parameter sets is acceptable within a series of data points with satisfactory behavior. They are referred to as the maximum left and right reaches. They can be summarized for all data points in a BReach plot (Figures 1b and 1c). For each data point on the x axis, the height of the gray area represents the distance between the maximum left reach (under the bisector) and right reach (above the

Figure 4. BReach analysis results for different synthetic data series, each represented by a horizontal bar. The gray tint indicates the geometry from which data points origin, characterized by its deviation in depth (D) and width (W) from the basic geometry. Dashed lines indicate detected consistent data periods. Diamonds under the bar indicate the resulting last data point of the period before transition. Diamonds above the bar indicate the resulting first data point of the period after transition.

When more relaxed degrees of tolerance are applied, results are allowed to be unacceptable in a higher percentage of the data points and thus only a limited subset of the data points within a reach is required to have acceptable model results. It is hence more likely that a subset exists that consists of comparable data points (i.e., data points that can be described by the examined parameter set). Hence, the influence of excesses in observational errors and the model's partial imperfection decreases and the maximum reach is able to pass a number of observations with nonacceptable model results. However, a significant change in geometry or flow conditions will interrupt this sequence of comparable observations. As a result, changes in data consistency will be emphasized in the corresponding BReach plot. Nevertheless, these consistency transitions can disappear in a BReach plot with a very high degree of tolerance.

Although they indicate changes in model performance precisely, BReach plots with low degrees of tolerance imply too wide a variety of possible causes for these changes. Plots with high degrees of tolerance on the other hand concentrate on the presence of changes in data consistency but the large failure tolerance prohibits locating these changes precisely. If BReach plots that combine all degrees of tolerance show consistent data periods, it is thus necessary to assess limits of consistent data periods based on all available results. Although user inference is required for this assessment, it is performed in a systematic manner. A first step is to decide on the amount of transition points to consider and to suggest a nearby index for each transition based on plots with high degrees of tolerance. If a clear discontinuity remains visible in these plots, it can be considered as a possible change in data consistency. A primary indication of the location of such a change is found by identifying the data points near this discontinuity that have a minimum distance between the bisector and the maximum left or right reach (red asterisks in the detail of Figure 1c).

bisector). In this plot, the vertical distance between the bisector and the maximum left reach indicates the maximum amount of data points before the investigated data point that can be described by at least one parameter set and under the prevailing degree of tolerance. For the data points after the index of interest, this maximum amount is indicated similarly by the vertical distance between the maximum right reach and the bisector.

This procedure is repeated for all degrees of tolerance. In Figures 2a–2c, BReach plots are combined for all applied degrees of tolerance, and thus summarize the results of an analysis.

2.1.6. Identification of Consistent Data Periods

Sudden changes in one of both reaches in a BReach plot are indications for changes in data behavior. In a combined BReach plot, all degrees of tolerance provide complementary information. As discussed in section 1, the stringent degrees of tolerance provide information on the diversity of the data points that is related with one of the following causes: (1) the occurrence of a higher observational error than estimated for the definition of acceptable results, (2) model deficiency in certain ranges of the investigated variable(s), and (3) data inconsistency.

Figure 5. Measured and adapted discharge at the upstream boundary of the MIKE 11 model [m^3/s].

ly constant value of this reach and on a nearly constant value of the maximum left reach for several data points that follow an indicative location (♦ in the detail of Figure 1b). The minimum of both indices indicates the first point of a new consistent period. With this approach, it is possible that the two resulting consistent data periods overlap and have some common data points. This is due to the fact that a transition in geometry or flow conditions is often a gradual change, and can hence cover several measurements that match both the initial and the new situation.

The example that is used to illustrate is a situation with a clear transition in data consistency. Albeit highly robust in all tested cases, it is not unthinkable that for some other situations the BReach methodology results in less distinctive plots and thus in a less evident decision process. Nevertheless, indications about data (in)consistency can be valuable to guide a more in-depth analysis that follows this preliminary data analysis.

2.2. Validation With Measured Data

The BReach methodology is tested on a stage-discharge data set measured in Aarschot. This measurement station has a well-documented history and thus allows for verifying changes in data consistency with information about historical events, changes in measuring device, and human interventions in the river. The station of Aarschot is located in the east of Flanders in the river Demer and has an upstream catchment of 2146 km^2 . The river Demer is a part of the Scheldt river basin (Figure 3). The river's regime is mainly influenced by rainfall events. In Aarschot, a small backwater effect occurs when stages in the river Dijle, in which the Demer discharges, are high. However, hydraulic model simulations indicate that this effect is smaller than the observational uncertainty that is assumed in section 2.1.3. Large infrastructure works to deepen and widen the river and to heighten the river dikes in the 1970's and the beginning of the 1980's had a large influence on the river's regime. Since then, the stage-discharge relation is known to be relatively stable. In Vanlierde [2013], an overview is given of the maintenance works between 2003 and 2010 in the proximity of Aarschot. The works that are expected to have an effect on the stage-discharge relation in Aarschot are a widening of the river bed (March–April 2008), the construction of a temporary dam that obstructed a

Based on these indicative location(s), BReach plots with more severe degrees of tolerance permit a final assessment of consistent data blocks. An example is illustrated in the detail of Figure 1b. If in these plots, the maximum right reach of several points that precede an indicative location obtains a nearly constant value, this value refers to a data point that is possibly the end of a first consistent period as no parameter set has a right reach that covers a larger subset of data points. Furthermore, a data point close to the indicative locations that precedes a sudden increase in the value of the maximum left reach after a series of data points with a nearly constant value for this reach also indicates a possible limit of a first consistent period. Both indices are indicated in the detail of Figure 1b (♦). Finally, the maximum of both indices is defined as the last data point of this first consistent period. Similarly, a starting point of a second consistent period can be extracted based both on the data point that coincides with a sudden increase in the value of the maximum right reach followed by a near-

part of the river's cross section (May–November 2007) and a reshaping of the winter bed profile (August 2009 to December 2010). These works were all located in a zone of 1 km upstream Aarschot.

For this data set, stages recorded manually during a discharge measurement are compared with stages recorded by the continuous measurement device. Stages that deviate more than 5 cm are treated as outliers and not used in the analysis. To prevent that a series of nearly simultaneous measurements dominates the results of the analysis, stage-discharge measurements performed within a time span of 2 h are averaged to result in one single data point if the difference in stage of these measurements does not exceed 3 cm. Hence, nearly simultaneous measurements are considered as separate information if the flow regime alters rapidly. Between 1980 and 2014, 277 stage-discharge data points are hence available in Aarschot. The stage of the lowest stage-discharge measurement is 14 cm higher than the lowest stage ever recorded and the stage of the highest measurement is 2 cm lower than the highest peak ever recorded.

2.3. Validation With Synthetic Data

A further validation of the BReach methodology is performed based on synthetic data created by use of hydraulic models. In previous studies, several authors already used results in internal nodes of one-dimensional hydraulic models for rating curve research. *Franchini et al.* [1999] used a variable parameter Muskingum-Cunge model to synthesize the rating curve in a cross section based on both a local stage record and a reliable rating curve in an upstream section of the same river. *Dottori et al.* [2009] used results of two hydraulic models as true data to test different rating curve formulas that account for unsteady flow conditions. Other studies use a hydraulic model for a sensitivity analysis to assess the contribution of one or more error sources to rating curve uncertainty [e.g., *Di Baldassarre and Montanari*, 2009; *Lang et al.*, 2010]. In

Di Baldassarre and Claps [2010] and *Neppel et al.* [2010], hydraulic models with different geometric situations are created based on historical information. Their impact on rating curves in an internal model node is investigated, with a particular focus on high flows. *Domeneghetti et al.* [2012] perturb discharge results of a hydraulic model with a Gaussian random error to create a synthetic stage-discharge data set. This data set is then used to assess rating curve uncertainty and its effect on the calibration of a second, different hydraulic model.

In this paper, several hydraulic models with different cross-section data are constructed as in *Di Baldassarre and Claps* [2010] and *Neppel et al.* [2010]. Each of these models is used to simulate a large range of stage situations and accompanying discharges that can be considered as a consistent data set. Results of different models can thus be combined to create synthetic data series with a known transition point between two consistent periods. Similar as in *Domeneghetti et al.* [2012], these data series are perturbed with a Gaussian noise that accounts for observational uncertainty. The consecutive steps of this approach are documented below.

Figure 6. (a) Basic geometry used for the validation case with limited model performance (gray) versus original geometry (black), both for the same location in the MIKE 11 model. (b) Synthetic rating curve data with a created abrupt change in data consistency at index 103 caused by a deepening of the river's cross sections with 1 m and a widening with 2 m (validation case with limited model performance).

Figure 7. Scatter plot with parameter values of the rating curve model in the x axis and (a) percentage of acceptable results in the y axis (b) distance to the maximum left reach in the y axis. Scatter plot with parameter values of the rating curve model both in the x and in the y axis for which the gray tint indicates the (c) percentage of acceptable results (d) distance to the maximum left reach.

2.3.1. Hydrodynamic Model

For the creation of the synthetic data, a MIKE 11 hydrodynamic model of the river Demer was available. This hydrodynamic model is constructed by the Flemish Hydrological Information Centre for operational water management purposes and was only slightly adapted for this research. The model covers the entire river reach as indicated in Figure 3b. The measurement station of Aarschot is situated internally in this model and is chosen as the location to create synthetic data.

The MIKE 11 software solves the continuity equation and the momentum equation of de Saint Venant using a six-point finite difference numerical scheme and thus enables to simulate a one-dimensional hydrodynamic flow problem in a river system [MIKE by DHI, 2011].

The river system is defined by a network of branches, nodes, and structures that represents the real geometric situation. The hydrodynamic problem is fully specified by boundary conditions at the network boundaries and by initial conditions. The model of the river Demer is implemented using fictitious river branches, representing the floodplains. These branches are connected to the river by spills, simulating the river dikes. For this research, the original model calibration based on the flood event of November 2010 [Bogman et al., 2013] is adopted without major changes. The only modification to the original model was the removal of the upstream part of the model, situated outside the region of interest.

The consequence of using a hydraulic model to create synthetic rating curve data is that model errors will affect these synthetic data. Therefore, it should be mentioned that although the roughness (Manning coefficient) in the hydraulic model of the river Demer varies spatially, it has a constant value for all stages of a cross section [Bogman et al., 2013]. However, effective Manning coefficients of a hydraulic model may vary for different flow conditions. Domeneghetti et al. [2012] show this for the river Po, using a model that solves the de Saint Venant equations with a four-point finite difference scheme (UNET). In this study, calibrated Manning coefficients decrease with increasing discharges, tending asymptotically to a constant value.

When the discharge exceeds the mean discharge (as reported in *Tarpanelli et al. [2013]*), this constant value is nearly reached. Despite of a slightly different (six-point) solution scheme, the model structure of MIKE 11 resembles the structure of the UNET model. Although the sizes of the river Po and the river Demer are not comparable, a further reflection on these results is important. Given these calibration results, using a constant roughness for all stages in the hydraulic model might result in a synthetic data set that fits the power law rating curve model (that also assumes this constant roughness) better than real stage-discharge data for low to medium flow situations. This issue will be further addressed in section 2.3.3.

2.3.2. Creation of a Transition in Data Consistency

A transition in data consistency is created by use of hydraulic models with different geometric properties. Based on the original model, 10 new models are created with altered cross sections over a zone of 2 km around Aarschot. Each of them deviates from the previous by lowering these cross sections with 0.1 m and widening them by 0.2 m, as visualized in the legend of Figure 4.

Next, a simulation that covers the whole range of observed water levels in Aarschot is performed with the original model and the 10 new models. Therefore, the measured discharges of the November 2010 flood are imposed at the upstream boundaries of the modeled tributaries. To ensure the coverage of the entire range of observed stages, the measured discharge at the upstream boundary in the river Demer itself is adapted (Figure 5). This adjustment is justified as the purpose of the simulations is to create synthetic data without needing a link to the real flood conditions.

For each simulation, the resulting stages and discharges in Aarschot form the basis of the synthetic data. To ensure a realistic variation in the synthetic data, the selection of points is based on the availability of in situ observed stage-discharge data in this location. Each gauging point of the measured data series is replaced

by the simulated water level that best approximates its stage value and by the accompanying discharge result. Hence, for all 11 geometric situations, a series of data points with simultaneous water level and discharge measurements is created.

The next step is to combine data originating from different geometric properties. Therefore, a randomly chosen index determines the moment of transition between two situations. Before this index, synthetic water levels and discharges created with the original geometry are used. After this transition index, the stage-discharge data result from simulations with a different geometry. The transition itself can be sharp or can consist of a number of data points originating from intermediate geometric situations. Finally, a Gaussian noise with zero mean and a standard deviation of 0.5 cm for water levels and of 3% for discharges was imposed on these data series to account for measurement errors.

By applying the BReach methodology for a comprehensive range of transition indices, different degrees of geometric changes and a variety of transition modes, its robustness can be tested. Figure 4 provides an overview

Figure 8. (a) Combined BReach plot for the gauging station in Aarschot in the river Demer. For each index in the x axis, the gray area indicates the span between the index of the maximum left reach (under the bisector) and of the maximum right reach (above the bisector). Each gray tint represents a different degree of tolerance. The red marks indicate the moments of the highest yearly maximum levels of the continuous measurement device. (b) Available stage-discharge data in Aarschot.

Figure 9. BReach analysis results for 20 subsamples of different synthetic data series, the latter each represented by a horizontal bar. The gray tint indicates the geometry from which data points origin, characterized by its deviation in depth (D) and width (W) from the basic geometry. For each of the 20 subsamples, diamonds under the bar indicate the detected last data point of the period before transition. Diamonds above the bar indicate their detected first data points of the period after transition. The more subsamples that share the same result, the larger the size of the diamond.

of the different situations used for validation, each represented by a horizontal bar. For each index in the data series, the gray tint indicates the geometric situation that is used to create it.

2.3.3. Effect of Data Availability

The BReach methodology can prove its robustness if results do not depend on a change in data availability. Synthetic data sets are created to investigate this effect.

First, a change in data availability is simulated by multiple random subsamples of the original synthetic data series. To achieve this, a synthetic data series is subsampled without replacement to form a new data series with only a quarter of the original synthetic data. Subsequently, a BReach analysis is performed on this reduced data set. This is repeated 20 times for data series 1, 4, 6, and 8 in Figure 4.

In a second part, a BReach analysis is performed with subsets of the data that contain only a limited range of discharge values. A low, medium, and high flow subset is created for data series 1, 4, 6, and 8 in Figure 4. The low flow subset consists of all gauging data corresponding with a discharge value lower than the mean discharge measured in Aarschot ($14.8 \text{ m}^3/\text{s}$). The high flow subset consists of all gauging data corresponding with a discharge value higher than the minimum of the measured annual maximum discharges ($30.6 \text{ m}^3/\text{s}$). The medium flow subset contains all gauging data that neither belong to the low nor to the high flow subset. This analysis provides insight in the capacity of measurements in different flow classes (each with their own parametrization results for the rating curve model) to distinguish between the created consistent periods and thus allows to estimate how a lack of (high or low) flow data might influence results. Furthermore, comparing results for low flow situations with those of the other two flow classes helps to understand the potential effect of using a fixed Manning coefficient for all stages in the hydraulic model (cf. section 2.3.1).

2.3.4. Effect of Observational Uncertainty

The estimation of the observational uncertainty is a major user-made decision for the cases applied in this research. In the definition of acceptable model results, an acceptance zone is assumed based on estimations of 95% uncertainty bounds for both stage and discharge measurements. As a Gaussian noise is imposed on the synthetic data series, these bounds consist of $\pm 2\sigma$, with σ the standard deviation of this noise. The

Figure 10. BReach analysis results for low, medium and high flow subsamples of different synthetic data series, the latter each represented by a horizontal bar. The gray tint indicates the geometry from which data points origin, characterized by its deviation in depth (D) and width (W) from the basic geometry. For each of the three subsamples, a colored diamond under the bar indicates the detected last data point of the period before transition. Diamonds above the bar indicate their detected first data points of the period after transition.

impact of this decision is investigated by changing these bounds for all synthetic data series of Figure 4 into $\pm\sigma$ and $\pm 3\sigma$ (68% and 99.7% uncertainty bounds).

Additionally, synthetic data sets are created similarly as data series 1 and 4 in Figure 4, but imposed to a Gaussian noise with a doubled standard deviation (1 cm for water levels and of 6% for discharges) and they are combined with the same acceptance zones as above. However, due to the altered standard deviations of the noise, these acceptance zones now consist of $\pm 0.5\sigma$, $\pm\sigma$, and $\pm 1.5\sigma$ (i.e., 38%, 68%, and 87% uncertainty bounds, respectively). This analysis thus expresses the effect of an underestimation of the observational uncertainty.

2.3.5. Effect of Model Deficiency

As stated above, the BReach methodology allows for a certain degree of deficiency in the model performance. Although in a part of the investigated river reach some inundations occur at higher stages, the majority of the rating curve data belong to a regular cross-section shape and can most likely be approached by a single power law. A test of the methodology with a less exemplary situation is thus necessary to demonstrate this flexibility toward model deficiency.

The original hydraulic model is adapted to create a sudden change in cross-section shape and consequently in the behavior of the rating curve. In the 2 km zone around Aarschot, where the maximum width of the original cross sections never exceeds 40 m, the part of each cross section situated more than 2.68 m above the lowest bed level is widened with 100 m. An example of this change is shown in Figure 6a. This adapted model is the starting point to create several synthetic data series with well-known transition points. These new data series can be expected to enclose two difficulties for the BReach analysis. First, there is no uniform rating curve behavior as a function of height and hence a simple power law cannot describe the full height range. Second, the difference between the data sets before and after the transition point is small for all data points above a water depth of 2.68 m (Figure 6b). Because of the large cross-section width above this water depth (up to 140 m), the geometric change between two analyzed situations (lowering with 1 meter and widening with 2 m) has only a small effect on the occurring discharges. In the created data series, 47%

of the data points are situated in this zone where the difference between simulation results (situation before and after geometric transition) is often smaller than the estimated observational uncertainty. The ability to utilize a model with known limitations to describe the data series is tested by using a simple power law model when applying the BReach methodology on those synthetic data series.

An additional BReach analysis is performed with low, medium, and high flow subsets for two of the synthetic data series investigated with the deficient rating curve model. These subsets are created similarly as in section 2.3.3. They should again provide insight in the capacity of measurements in different flow classes to distinguish between the created consistent periods, this time with a particular focus on the (expected) limited contribution of high flow data.

3. Results and Discussion

3.1. Parametrization

Although the aim of the BReach methodology is not to assess the parameters of a rating curve model, a verification of the sampling strategy is important. As all three parameters are sampled from truncated distributions, the validity of the chosen sampling limits needs to be evaluated. An example of such an evaluation is shown in Figure 7 for data series 3 of Figure 4. For each sampled parameter set, the percentage of points in this data series that has acceptable results is assessed. This percentage indicates the overall performance of the parameter set for the complete data series. Figure 7a shows this percentage in the y axis of three scatter plots, with in the x axis the corresponding value of one of the rating curve parameters. For parameters n and c , it is clear that the sampled range is wider than the zone where maximum values of this percentage occur. For parameter h_0 however, percentages at the lower limit of the sampled zone are only slightly lower than the maximum attained percentages.

In a second analysis, parameter values that are determinant for a BReach analysis are evaluated. Therefore, for each point in the synthetic data series, the distance to the maximum left reach is defined as the difference

Figure 11. Combined BReach plots of a synthetic data series with a created abrupt change in data consistency at index 201. The analyzed data set is a subsample that consists of (a) low flow data, (b) medium flow data, (c) high flow data. In all subplots, the gray area indicates for each index in the x axis the span between the index of the maximum left reach (under the bisector) and the maximum right reach (above the bisector). Each gray tint represents a different degree of tolerance.

ence between the index of the analyzed data point and its maximum left reach. Moreover, the parameter sets that cause this maximum left reach are selected. The three scatter plots in Figure 7b show the distance to the maximum left reach in the y axis, with in the x axis the corresponding values of one of the rating curve parameters. As the maximum reach in a data point is caused by only a limited amount of parameter sets, this plot shows a very irregular spread in the scatter of points. The zones with the highest distances to the maximum left reach are rather similar to the zones with maximum percentages in Figure 7a, albeit a bit more constrained for parameter h_0 .

In Figure 7c, scatter plots show two dimensions of the three-dimensional parameter space. The gray tint indicates the percentage of acceptable results for the corresponding parameter set. In Figure 7d, similar plots are shown with a gray tint representing the distance to the maximum left reach. They have the same irregular spread in the scatter of points as the plots of Figure 7b. In both Figure 7c

and 7d, a strong interdependency of all three parameters is visible. Again, similar patterns are detectable for both the percentages of acceptable results and the distances to the maximum left reach. The plots with parameter h_0 in the x axis show two separate groups of parameters (situated parallel to each other) that represent the two different consistent periods in the data series.

Based on this parameter analysis, the lower limit of the sampling range of parameter h_0 (the stage of zero flow) could be questioned. Nevertheless, this limit is based on physical properties of the river (40 cm lower than the river bed of the hydraulic model). As discussed in section 2.1.2, it is thus not advisable to lower it. Therefore, the original sampling strategy is maintained.

3.2. Measured Data Set

In Figure 8a, a combined BReach plot for Aarschot is shown. In this plot, the moments near the five highest yearly maximum stages recorded with the continuous measurement device are indicated with a red mark. This plot shows changes in data consistency that corresponds with historical information. During a first period (February 1980 to May 1982), there is no consistency detected in the data series. This period coincides with the infrastructural works to deepen and widen the river's cross section and to heighten the river dikes, resulting in a continuous change in the stage-discharge relationship. The end of these works results in the start of a consistent period at index 29 (August 1982) that ends at index 235 (March 2006). After this moment, the data series shows a lack of consistency again. The moment of this second change is situated between two large floods and near the introduction of new measurement devices. Since that moment, these new devices that make use of the Doppler effect gradually replaced the former mechanical velocity measurements. Moreover, the maintenance works that effect the cross section of the river bed occurred in this period. It is difficult to quantify the effect of these causes separately. Nevertheless, it is possible that their joined effect leads to the detected change in consistency. In Figure 8b, stage-discharge measurements of these three different periods are plotted separately.

Figure 12. Combined BReach plots of synthetic data series with a created abrupt change in data consistency at index 201 and imposed to a noise with a standard deviation of 0.5 cm for stages and 3% for discharges. The zone that defines acceptable model results encloses an uncertainty bound around stage and discharge measurements of (a) 68%, (b) 95%, (c) 99.7%. In all subplots, the gray area indicates for each index in the x axis the span between the index of the maximum left reach (under the bisector) and the maximum right reach (above the bisector). Each gray tint represents a different degree of tolerance.

3.3. Basic Synthetic Data Sets

For all synthetic data series, the end-point of the first and the startpoint of the second consistent data period are derived according to step 6 of the BReach methodology (section 2.1.6). In Figure 4, these points are represented by filled diamonds.

Data series 1–5 represent an abrupt transition from the original geometry toward a 1 m deepened and 2 m widened geometry. In Figures 2a–2c, combined BReach plots, respectively, belonging to data series 4, 1, and 2 are shown. These plots demonstrate how data points that systematically cause discontinuities for a 0% degree of tolerance evolve toward singularities (5% and 10% degrees of tolerance) that are at last concealed by the flexibility of the 20% and 40% degrees of tolerance. In these figures, a discontinuity in the value of the maximum reaches is visible for all degrees of tolerance, indicating a transition in data consistency. These transitions coincide with the created transitions in the synthetic data series. For data series 3 and 5, these plots are likewise self-explanatory and

Figure 13. BReach plots of synthetic data series with a created abrupt change in data consistency at index 201 and imposed to a noise with a standard deviation of 1 cm for stages and 6% for discharges. The zone that defines acceptable model results encloses an uncertainty bound around stage and discharge measurements of (a) 38%, (b) 68%, (c) 86.6%. In all subplots, the gray area indicates for each index in the x axis the span between the index of the maximum left reach (under the bisector) and the maximum right reach (above the bisector). Each gray tint represents a different degree of tolerance.

Figure 9. Due to the random character of the subsamples, the real transition point is not always included in the investigated subsamples and the most neighboring data point in the subsample will hence represent the transition moment. For some of the subsamples, the randomly picked data points lead to less straightforward BReach plots. Nevertheless, a detection of the correct or neighboring indices was possible for 95% of the subsamples of synthetic data series 1, while 70% were found for data series 2. In all other cases, other data points near the original transition point were detected (Figure 9). In one of the subsamples of data series 1, an extra transition moment was detectable in the BReach plots. This false transition point has no harmful consequences for the following in-depth analysis within the consistent data periods that are derived from the BReach results. Nevertheless, the data series that are used for it will be unnecessarily shortened.

When a more gradual transition is considered (data series 3), all subsamples result in detecting both indices in the transition zone. In the analysis for a smaller change in geometry (data series 4), 60% of all detected indices equal or neighbor the transition point.

Despite the sometimes larger scatter, Figure 9 clearly indicates that the detected changes in data consistency all pinpoint at the period in which a real transition occurs. For all investigated cases, the BReach methodology is thus capable of delimiting a change in consistency irrespectively of a more limited data availability.

In Figure 10, results of the BReach analysis with low, medium, and high flow subsamples are shown. In all cases, it was possible to detect changes in consistency near the created transition point of the synthetic data series. For the data series with only 0.5 m deepening and 1 m widening of the cross sections however, an extra transition point was detectable in the BReach plots. Although all subsamples lead to nearly the same conclusions, results tend to be less straightforward when discharge values are higher (e.g., Figure 11). For these higher discharges, the relative contribution of the geometric changes between two consistent

do not cause difficulties when assessing consistent data periods. The BReach methodology leads thus to a correct assessment of the transition points for these abrupt transitions.

When a more gradual transition is simulated (data series 6 and 7), consistent data periods are all found within or at the boundaries of this transition zone. For data series 6, the combined BReach plots result in two separate consistent data periods, divided by several consecutive data points that do not belong to either of them.

A less significant geometric transition is tested with data series 8 and 9. A 50 cm deepening and 1 m widening of the cross sections is still clearly detectable. When applying a 10 cm deepening and a 20 cm widening however, the deviations of the rating curve caused by geometric changes have the same order of magnitude as the estimated data uncertainty and no clear transitions are detected.

3.4. Effect of Data Availability

The results of the BReach analysis for 20 subsamples that contain a quarter of the original data series are shown in

Figure 14. (a) BReach analysis results based on a deficient model for different synthetic data series, each represented by a horizontal bar. The gray tint indicates the geometry from which data points origin, characterized by its deviation in depth (D) and width (W) from the basic geometry. Dashed lines indicate detected consistent data periods. Diamonds under the bar indicate the resulting last data point of the period before transition. Diamonds above the bar indicate resulting first data points of the period after transition. (b) Combined BReach plots of synthetic data series 1 with a created abrupt change in data consistency at index 48. For each index in the x axis, the gray area indicates the span between the index of the maximum left reach (under the bisector) and the maximum right reach (above the bisector). Each gray tint represents a different degree of tolerance.

for a 0% degree of tolerance (Figure 12c). However, the reduction of type II errors, that causes this change, risks to be accompanied by an increase in type I errors (false acceptances) that can lower the discriminatory capacity of the methodology.

In Figures 13a–13c, results are shown of using the same acceptance zones for a synthetic data series that is imposed to a Gaussian noise with doubled standard deviation. This leads to effective acceptance bounds of, respectively, $\pm 0.5\sigma$ (38%), $\pm \sigma$ (68%), and $\pm 1.5\sigma$ (86.6%). An acceptance zone of $\pm 0.5\sigma$ is too small and results in excessive amounts of type II errors and a corresponding combined BReach plot that does not provide useful information. Nevertheless, both other figures indicate a transition around index 201.

Similar results were obtained for all other data series in the analysis. Although a correct estimation of observational uncertainties should be aimed for, a (limited) misjudgment does not seem to alter results of a BReach analysis. Moreover, the information content of BReach results based on multiple acceptance zones has a complementary character and can augment the comprehension of results.

3.6. Effect of Model Deficiency

Despite the use of a model that is known to cover only a limited range of the measured stages, all investigated situations lead to the assessment of indices near the real data transition point (Figure 14a). Figure

periods to the channel conveyance is lower and thus the distinctive capacity of the corresponding subsample decreases. For the analyzed data sets, a lack of high flow data would thus not influence the results of a BReach analysis strongly. However, the large distinctive capacity of the low flow subsample is possibly too optimistic due to the fixed Manning coefficient for all stages in the hydraulic model (cf. discussion in section 2.3.1). It is important to note that these results are situation-specific and depend on the type of events that cause temporal consistency changes and of their characteristics.

3.5. Effect of Observational Uncertainty

A change in the size of the acceptance zone that defines acceptable model result alters the structure of the BReach plots. This is illustrated in Figure 12. As can be expected, a smaller acceptance zone of $\pm \sigma$ (68% uncertainty bounds) will result in more false rejections (type II errors) and thus BReach plots with low degrees of tolerance will not contain much information (cf. Figure 12a). However, despite a lot of singularities in the plots with higher degrees of tolerance, the pattern of the data consistency is still detectable. An increased acceptance zone of $\pm 3\sigma$ (99.7% uncertainty bounds) on the contrary, results in a very straightforward BReach plot, even

Figure 15. Combined BReach plots based on a deficient model for a synthetic data series with a created abrupt change in data consistency at index 201. The analyzed data set is a subsample that consists of (a) low flow data, (b) medium flow data, (c) high flow data. In all subplots, the gray area indicates for each index in the x axis the span between the index of the maximum left reach (under the bisector) and the maximum right reach (above the bisector). Each gray tint represents a different degree of tolerance.

14b shows the combined BReach plot for synthetic data series 1 as an example. Similarly as in the previous section, plots for low degrees of tolerance contain little information while plots for higher degrees lead to a correct assessment of the transition point.

An evaluation of data series 1 and 4 of Figure 14 using subsamples with only low, medium, and high flow data leads to satisfying results for the low and medium flow data. As could be expected from Figure 6b, a BReach analysis of only the high flow data does not lead to the assessment of consistent periods (e.g., Figure 15). Again, these results are situation-specific and cannot be extrapolated to other situations without knowledge of the governing circumstances.

4. Conclusions

In this paper, a methodology is developed to assess consistency in rating curve data preliminary to a more profound model evaluation and it is successfully validated with both measured and synthetic sets of rating

curve data. The combination of different degrees of tolerance results in a robust methodology and limits the amount of assumptions to be made by users.

Tests show relatively little dependency of the results on possible misjudgments. Even the choice of a model that is only able to describe a subset of the analyzed data leads to satisfying conclusions. Nevertheless, a preliminary BReach analysis should be followed by a more in-depth analysis within the identified consistent data periods.

Due to its robustness and to its possibility to extend an existing analysis with new data without demanding calculations, the BReach methodology is a very promising tool for preliminary data analysis in both a research and an operational context. Its flexibility to adopt different approaches for sampling and for the definition of acceptable results permits a relatively simple addition of BReach within an existing context for data and model analysis. Moreover, the use can be expanded. A test with a larger diversity of rating curve data and an application for other processes with temporal or seasonal variability are interesting future research steps. Furthermore, the methodology can be used in other dimensions than the temporal one. For example, performing an additional BReach analysis on data points that are sorted as a function of stage might provide information about stage-dependent discontinuities and thus the need to use segmented rating curves. Besides, an extension toward other variables and processes can be investigated.

References

- Beven, K. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, 320(1–2), 18–36, doi:10.1016/j.jhydrol.2005.07.007.
- Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, 6, 279–298, doi:10.1002/hyp.3360060305.
- Blazkova, S., and K. Beven (2009), A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, *Water Resour. Res.*, 45, W00B16, doi:10.1029/2007WR006726.
- Bogman, P., S. Buitrago, L. Coen, S. Smets, F. Pereira, P. Peeters, T. Verwaest, and F. Mostaert (2013), Modelactualisatie en opmaak overstromingskaarten ihkv. ROR, *Tech. Rep. 00_154*, Flanders Hydraul. Res., Antwerp, Belgium.

Acknowledgments

The BReach code is available from the authors upon request (katrien.vaneerdenbrugh@ugent.be). The rating curve data of Aarschot were obtained freely from the Flemish Hydrological Information Centre (hic@vlaanderen.be). The authors thank Flanders Hydraulics Research and its Hydrological Information Centre for providing these gauging data and accompanying information, access to MIKE 11 software and the model of the river Demer and Waterwegen en Zeekanaal NV for providing information about infrastructure works in Aarschot. Furthermore, they thank Henrik Madsen for providing additional access to MIKE 11 software, Bernard De Baets and Keith Beven for their highly appreciated feedback and Giuliano Di Baldassarre and the two anonymous reviewers for their constructive comments and helpful suggestions on an earlier version of the manuscript.

- Di Baldassarre, G., and A. Montanari (2009), Uncertainty in river discharge observations: A quantitative analysis, *Hydrol. Earth Syst. Sci.*, 13(6), 913–921, doi:10.5194/hess-13-913-2009.
- Di Baldassarre, G., and P. Claps (2010), A hydraulic study on the applicability of flood rating curves, *Hydrol. Res.*, 42(1), 10–19, doi:10.2166/nh.2010.098.
- Domeneghetti, A., A. Castellarin, and A. Brath (2012), Assessing rating-curve uncertainty and its effects on hydraulic model calibration, *Hydrol. Earth Syst. Sci.*, 16(4), 1191–1202, doi:10.5194/hess-16-1191-2012.
- Dottori, F., M. L. V. Martina, and E. Todini (2009), A dynamic rating curve approach to indirect discharge measurement, *Hydrol. Earth Syst. Sci.*, 13(6), 847–863, doi:10.5194/hess-13-847-2009.
- Franchini, M., P. Lamberti, and P. Di Giammarco (1999), Rating curve estimation using local stages, upstream discharge data and a simplified hydraulic model, *Hydrol. Earth Syst. Sci.*, 3(4), 541–548, doi:10.5194/hess-3-541-1999.
- Jalbert, J., T. Mathevet, and A.-C. Favre (2011), Temporal uncertainty estimation of discharges from rating curves using a variographic analysis, *J. Hydrol.*, 397(1–2), 83–92, doi:10.1016/j.jhydrol.2010.11.031.
- Krueger, T., J. Freer, J. N. Quinton, C. J. A. Macleod, G. S. Bilotta, R. E. Brazier, P. Butler, and P. M. Haygarth (2010), Ensemble evaluation of hydrological model hypotheses, *Water Resour. Res.*, 46, W07516, doi:10.1029/2009WR007845.
- Lang, M., K. Pobanz, B. Renard, E. Renouf, and E. Sauquet (2010), Extrapolation of rating curves by hydraulic modelling, with application to flood frequency analysis, *Hydrol. Sci. J.*, 55(6), 883–898, doi:10.1080/02626667.2010.504186.
- Le Coz, J., B. Renard, L. Bonnifait, F. Branger, and R. Le Boursicaud (2014), Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach, *J. Hydrol.*, 509, 573–587, doi:10.1016/j.jhydrol.2013.11.016.
- McMillan, H. K., and I. K. Westerberg (2015), Rating curve estimation under epistemic uncertainty, *Hydrol. Processes*, 29, 1873–1882, doi:10.1002/hyp.10419.
- McMillan, H., J. Freer, F. Pappenberger, T. Krueger, and M. Clark (2010), Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions, *Hydrol. Processes*, 24, 1270–1284, doi:10.1002/hyp.7587.
- McMillan, H., T. Krueger, and J. Freer (2012), Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality, *Hydrol. Processes*, 26, 4078–4111, doi:10.1002/hyp.9384.
- MIKE by DHI (2011), MIKE 11, A modelling system for rivers and channels, Reference manual, DHI, Hørsholm, Denmark.
- Morlot, T., C. Perret, A. Favre, and J. Jalbert (2014), Dynamic rating curve assessment for hydrometric stations and computation of the associated uncertainties: Quality and station management indicators, *J. Hydrol.*, 517, 173–186, doi:10.1016/j.jhydrol.2014.05.007.
- Moyeed, R., and R. Clarke (2005), The use of Bayesian methods for fitting rating curves, with case studies, *Adv. Water Resour.*, 28(8), 807–818, doi:10.1016/j.advwatres.2005.02.005.
- Neppel, L., B. Renard, M. Lang, P.-A. Ayrat, D. Coeur, E. Gaume, N. Jacob, O. Payrastre, K. Pobanz, and F. Vinet (2010), Flood frequency analysis using historical data: Accounting for random and systematic errors, *Hydrol. Sci. J.*, 55(2), 192–208, doi:10.1080/02626660903546092.
- Pappenberger, F., P. Matgen, K. J. Beven, J.-B. Henry, L. Pfister, and P. de Fraipont (2006), Influence of uncertain boundary conditions and model structure on flood inundation predictions, *Adv. Water Resour.*, 29(10), 1430–1449, doi:10.1016/j.advwatres.2005.11.012.
- Pelletier, P. M. (1988), Uncertainties in the single determination of river discharge: A literature review, *Can. J. Civ. Eng.*, 15(5), 834–850.
- Petersen-Øverleir, A. (2006), Modelling stage-discharge relationships affected by hysteresis using the Jones formula and nonlinear regression, *Hydrol. Sci. J.*, 51(3), 365–388, doi:10.1623/hysj.51.3.365.
- Petersen-Øverleir, A., and T. Reitan (2005), Objective segmentation in compound rating curves, *J. Hydrol.*, 311(1–4), 188–201, doi:10.1016/j.jhydrol.2005.01.016.
- Petersen-Øverleir, A., and T. Reitan (2009), Bayesian analysis of stage fall discharge models for gauging stations affected by variable backwater, *Hydrol. Processes*, 24, 3057–3074, doi:10.1002/hyp.7417.
- Reitan, T., and A. Petersen-Øverleir (2011), Dynamic rating curve assessment in unstable rivers using Ornstein-Uhlenbeck processes, *Water Resour. Res.*, 47, W02524, doi:10.1029/2010WR009504.
- Schmidt, A. R., and B. C. Yen (2001), Stage-discharge relationship in open channels, *Proceedings of the 3rd International Symposium on Environmental Hydraulics*, edited by D. Boyer and R. Rankin, ISEH and IAHR, Tempe, Ariz.
- Tarpanelli, A., S. Barbetta, L. Brocca, and T. Moramarco (2013), River discharge estimation by using altimetry data and simplified flood routing modeling, *Remote Sens.*, 5, 4145–4162, doi:10.3390/rs5094145.
- Vanlierde, E. (2013), Sediment concentrations, fluxes and source apportionment: Methodology assessment and application in Nete and Demer tributary basins (river Scheldt basin, Belgium), PhD thesis, Dep. of Geol. and Soil Sci., Ghent Univ., Ghent, Belgium.
- Van Hoey, S., I. Nopens, J. van der Kwast, and P. Seuntjens (2014), Dynamic identifiability analysis based model structure evaluation considering rating curve uncertainty, *J. Hydrol. Eng.*, 10(5), 04014072, doi:10.1061/(ASCE)HE.1943-5584.0000995.
- Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Processes*, 17(2), 455–476, doi:10.1002/hyp.1135.
- Westerberg, I., J.-L. Guerrero, J. Seibert, K. J. Beven, and S. Halldin (2011), Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras, *Hydrol. Processes*, 25(4), 603–613, doi:10.1002/hyp.7848.